Available online at www.elixirpublishers.com (Elixir International Journal)

Information Technology

Elixir Inform. Tech. 129 (2019) 53011-53013

ISSN: 2229-712X

Machine Learning Algorithms and Applications: A Review

Foteini Emmanouilidou
IT Department, My Company Projects O.E., Thessaloniki, 54622, Greece,

ARTICLE INFO

Article history:
Received: 5 March 2019;
Received in revised form:
1 April 2019;
Accepted: 8 April 2019;

Keywords
Machine Learning,
Artificial Intelligence,
Algorithms,
Applications,
Data,

ABSTRACT
This paper will discuss about Machine Learning technology, its algorithms and its applications. In the new era of automations, business intelligence and the growing datacenters capacity, machine learning is a key component. Through literature review this paper demonstrates this new field and its characteristics. Different algorithms exist for various types of works, making it possible for average business to use the one that suits its needs the best. Machine Learning has come to stay for many years, providing added value to data and predictive mechanisms that will make our world a better place.

1. Introduction

Machine learning is the application of artificial intelligence (AI) which gives systems the ability to assimilate and automatically improve from experience without being expressively programmed. The focus of machine learning is basically the advancement of computer programs with the ability to apply the data accessed for its advancement. Data or observation begins the learning process. For example, direct familiarity or direction that will enable it to recall data pattern thereby making an improved decision in the future using the provided examples as a guide. The primary objective is to give computers the ability to automatically learn and take decisions accordingly without human interference or guide. This paper is organized as following: Section 2 gives a clear categorization of Machine Learning Algorithms' categories. Section 3 describes the most common Machine Learning Algorithms and some applications of theirs. Conclusion of work is presented in section 4.

2. Categorization of machine learning algorithms

Machine learning algorithms can be divided into four which are reinforcement, unsupervised, semi-supervised and supervised, as shown in figure 1.

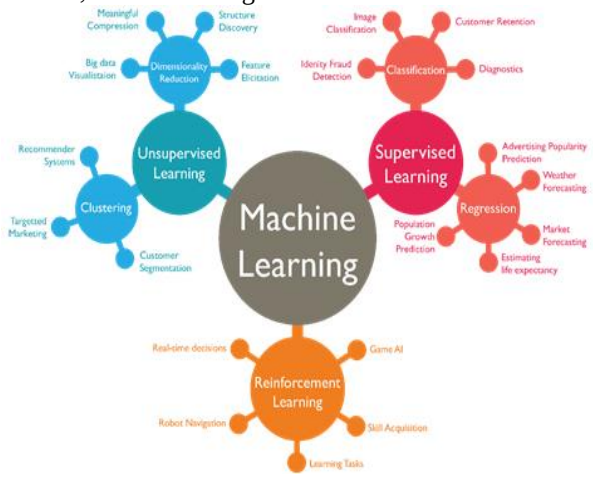

Figure 1. Machine Learning Categorization.

Supervised Learning

For supervised learning, examples are used to teach the machine. An operator stores the desired inputs and outputs into a dataset that is provided to the machine learning algorithm to enable it to arrive at the desired outcome. The algorithm recognizes the data patterns and learns from them through experience and making predictions which are corrected by the operator who already knows the answers to the problem. The main objective is for the algorithm to attain a high level of performance and efficiency creating a continuous process until it is achieved.

The subcategories under supervised learning are classification, regression, and forecasting.

Classification: The machine learning algorithm has to make a conclusion from the observations gotten from values and be able to determine the category of that observation. For example, during email filtering into “spam” and “not spam”, available data will be observed by the program to recognize the input and at such accordingly filter them.

Regression: The machine learning algorithm understands and makes evaluations from the connections that exist among variables.

In regression tasks, much focus is placed on an independent variable and a couple of other dependent fluctuating variables which gives it the ability to anticipate and predict future occurrence.

Forecasting: This is the ability to calculate or predict some future event or condition usually as a result of study and analysis of available data used commonly for the analysis of trends.

Semi-supervised learning

The supervised and semi-supervised learning are quite similar but the semi-supervised makes use of unlabeled and labeled data. Labeled data is one that contains useful information with proper documentation for easy comprehension by the algorithm while unlabeled data do not have this useful information but allows the algorithm the

Tele:
E-mail address: foteini@mycompany.com.gr

ability to experience and understand unlabeled data giving it the ability to label them.

Unsupervised learning

In unsupervised learning, the machine learning algorithm is able to identify patterns by studying available data. The machine learning algorithm analyses the available data and finds the relationship or correlation between them without human interference or input data. In the process of unsupervised learning, the machine learning algorithm without supervision analyses sets of large data and filters them accordingly. The algorithm through the data structure tries to organize the data or groups them in a cluster.

This enables the algorithm to make decisions, get improved through access to more data thereby making it more exceptional.

Sub categories under unsupervised learning are:

Clustering: This process involves using a set of defined criteria for the grouping of data that are similar. It is very useful when creating data groups for easy analysis in order to find correlations or patterns.

Dimension reduction: In this process, the required information is gotten from the reduction of the number of variables under consideration.

Reinforcement Learning

Reinforcement learning centers around a learning process that is regimented where a set of rules, action and data value are provided to the machine algorithm. These rules enable the algorithm to investigate different options and conceivable outcomes by observing and assessing each outcome to figure out the most ideal. Reinforcement learning teaches machine experimentation. It gains from previous experiences and then starts to adjust its methodology in light of the circumstance to accomplish the most ideal outcome.

3. Algorithms and Applications

Machine Learning works in a very specific way, as shown in figure 2.

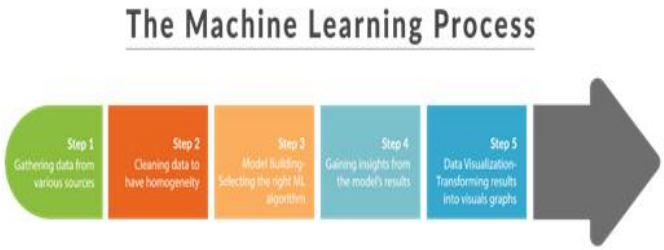

Figure 2. Machine Learning Process.

The most common and popular machine learning algorithms that are used according to references are listed below:

Naïve Bayes Classifier Algorithm (Supervised Learning - Classification)

The Naïve Bayes Classifier depends on the Bayes theorem which categorizes every value free from other values. It enables class/category predictions by the use of probability based on a set of given features.

K Means Clustering Algorithm (Unsupervised Learning - Clustering)

This is a kind of unsupervised learning algorithm whose main objective is the categorization of unlabeled data. i.e. data absent of groups or categories. The functions of the algorithm lie in the search for groups inside the data having connections with the K variable representing that number of groups. It then utilizes the features provided to iteratively assign every data point to the k groups.

Support Vector Machine Algorithm (Supervised Learning - Classification)

This is a supervised learning algorithm used in the analysis of data meant for regression and classification analysis. By making available sets of examples to be used for training, they are able to categorize data where each marked set is a property of one of the two available categories. The algorithm afterward designs a model responsible for attaching value to any of the categories.

Linear Regression (Supervised Learning/Regression)

This is primarily the type of regression that enables us to know the connection between two variables that are continuous.

Logistic Regression (Supervised learning – Classification)

This centers around evaluating the likelihood of an event to occur using a previously provided data as a basis. It is used in covering variables that depends on binary, i.e. where only two values exist and represent the outcome.

Artificial Neural Networks (Reinforcement Learning)

Artificial Neural Network (ANN) consists of a series of layers known as 'units' which have connecting layers on each side. Artificial Neural Networks are biologically inspired for example, how the brain processes information. ANNs is a housing of interconnected processing elements having a single goal at finding a solution to a specific task.

ANNs also assimilate by experience and example and they also have the ability to use high-dimensional data to model non-linear relationships where there is difficulty in understanding the connections amongst the input variables.

Decision Trees (Supervised Learning–Classification/ Regression)

A decision tree is a structure that is similar to a flow-chart that utilizes branches in illustrating the result of every possible decision. Every node on the tree is a representation of a test carried out on a particular variable that will result in a branch that represents that test's outcome.

4. Conclusion

AI and especially Machine Learning are exciting new concepts that have arrived to change our lives. Though the huge potential of such concepts is already visible, lack of compute power that currently exists in our infrastructures, security issues that many arise out of the "training" on sensitive data the lack of highly trained AI professionals, are only some of the current and future challenges that need to be addressed.

References

R. Feldman and J. Sanger (2007). The text mining handbook: advanced approaches in analyzing unstructured data. Imagine, vol. 34, pp. 410.

E. Guzman, M. El-Haliby, and B. Bruegge (2015). Ensemble Methods for App Review Classification: An Approach for Software Evolution (N). 30th IEEE/ACM International Conference on Automated Software Engineering, 771–776.

Chandan Prasad Gupta, BalKrishan (2015). Detecting Sentiment Analysis in Nepali Texts, IEEE, 1-4.

Zhongkai Hu, Jianqing Hu,Weifeng Ding,Xiaolin Zheng (2015). Review Sentiment Analysis based on Deep Learning. 12th International Conference on ebusiness Engineering, IEEE, 87-94.

Andrea Salinca (2015). Business Review Classification employing Sentiment Analysis. 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, 247-250.

Nagamma P, Pruthvi H.R, Nisha K.K and Shweta NH (2015). An Improved Sentiment Analysis of Online Movie Review

Based on Clustering for Box Office Prediction. International Conference on Computing Communication and Automation, 933- 937.

Mrs.J. Sukanya (2017). Applications of Big Data Analytics and Machine Learning Techniques in Health Care Sectors. International Journal of Engineering and Computer Science, vol.6, 21963-21967.

M. Senturk (2014). Breast cancer Diagnosis via Data Mining: Performance Analysis of Seven Different Algorithms. Computer Science & Engineering", vol.1, 1-10.

E. Venkatesan (2015). Performance analysis of decision tree algorithms for breast cancer classification. Indian Journal of Science and Technology, vol.8, 1-8.

https://medium.com/marketing-and-entrepreneurship/10-companies-using-machine-learning-in-cool-ways-887c25 f913 c3

https://pythontips.com/2017/11/11/introduction-to-machine-learning-and-its-usage-in-remote-sensing/